194-00312TPW

User Characterization and Requirements Analysis

White Paper

Working paper - Not intended for formal review or Government approval.

September 1994

Prepared Under Contract NAS5-60000

RESPONSIBLE SCIENTIST

Tess Wingo /s/ 9/30/94

Tess Wingo, Science Specialist, ESSi
EOSDIS Core System Project

SUBMITTED BY

Robert J. Curran /s/ 9/30/94

Robert J. Curran
EOSDIS Core System Project

Hughes Applied Information Systems
Landover, Maryland

This page intentionally left blank.

Contents

1. Introduction

2. ECS User Community

3. System Access

4. Data Extraction and Distribution

5. Observations and Implications

6. Current Status of Analysis

Figures

Tables

Abbreviations and Acronyms

1. Introduction

1.1 Purpose

The purpose of this Technical Paper is to present (as of the Second Quarter 1994) the ECS Project's characterization of the projected EOSDIS user communities for the data products from the AM1, TRMM (LIS and CERES) and SeaWiFS sensors in the 1999-2003 time frame, and to summarize how these users will interact with the system. In order to focus future refinements, this paper is meant to “bound the problem” by identifying a range of numbers to support design decisions and design drivers. The methodology for much of the information presented in this paper is explained in Technical Paper # 194-00313TPW, “User Characterization Methodology and Results”. The science scenario matrix purpose and development, scenario collection and analysis, demographics, functional analysis, and high level characterization of the community are clarified in that document. The collected user scenarios and associated demographic information can be found in the “User Scenario Notebook,” # 194-00311TPW.

1.2 Organization

This document is organized into five sections:

Section one attempts to provide a maximum and minumum boundary for the number of users in the EOS Science, General Science, and Non-Science ECS user communities. The second section provides an understanding of the frequency of accesses to the system; the method of access, whether direct or indirect through other systems; and the access paths through various service components. The third section deals with the volumes of data extracted and distributed, the methods of search requests for data to be extracted and distributed, and the media for data distribution. Section four restates many of the design observations and implications. Finally, the fifth section evaluates the current status of analysis.

1.3 Review and Approval

This White Paper is an informal document for internal (i.e. Project) distribution approved at the Office Manager level. It does not require formal Government review or approval; however, it is submitted with the intent that review and comments will be forthcoming.

The material in this paper was used to supply “order of magnitude” bounds on potential user communities, and was for SDR use only to indicate issues affecting overall system design. Subject to large error bars these estimates will be refined for PDR based on projections and lessons learned from existing systems. When new or modified characteristics of user interaction based upon early experience with Version 0, the projected migration of Version 0 data products, and more in-depth analysis of user service access rates become available these ideas will be validated .

In several sections of this paper reference is made to subjective judgments based upon "prior experience" and "experience with previous remotely sensed research products". These references are to the experience of the contributors to this document in working with Federal data centers, Federal agencies, bureaus and laboratories, state agencies and commerical organizations in there application of remotely sensed data. Questions regarding these judgments and other questions regarding technical information contained within this Paper should be addressed to the following ECS contacts:

- ECS Contacts
  - Tess Wingo, Science Specialist, ESSi, (301) 925-0814, twingo@eos.hitc.com
  - Pitt Thome, ESSi, (301) 925-0807, pthome@eos.hitc.com
  - Celeste Jarvis, Program Manager, ESSi, (301) 925-0800, cjarvis@eos.hitc.com

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems, Inc.
1616A McCormick Dr.
Landover, MD 20785

2. ECS User Community

2.1 Potential Number of Users

It is important to understand the number of potential users for many reasons. For example, the number of users influences many design areas including services requested, system load, number of user services personnel, needed response times, and the potential number of concurrent users of the system.

In the highest level of analysis, users were placed into three categories: EOS Science, General Science, and Non-Science. Different methods were used to assess the size of each category. This paper attempts to provide maximum and minumum boundaries for the number of ECS users. Future analysis will help to narrow the range within those bounds.

Estimating the number of users is important because it allows us to characterize system load in terms of accesses and complexity of accesses. It also allows us to size services to users. Large numbers indicate that ECS will not be able to provide all services to all users. Therefore, ECS must be looking for additional service providers (shown in Section 2.1.3).

The “number of users” was obtained using a variety of methods. The number of science users was derived from Peterson’s Guide to Graduate Programs in the Physical Sciences and Mathematics, and from an analysis done on various earth science journals. Users classified as non-science users were from Federal Government, states, commercial, education, library, and policy maker organizations. This paper also documents the various methods used to obtain the number of users for EOSDIS from each of these community sectors.

2.1.1 Science Users

Science users are important because the mission of NASA’s EOS program is to support interdisciplinary research and model development, which will help scientists detect changes, understand the processes that control our global environment, improve predictions of events, and comprehend the consequences of human activities.

Table 2-1 show the science communities and their estimated sizes. Science users include EOS Science and General (Non EOS-funded) Science. Within these categories various estimation methods, sources and assumptions were used.

EOS-funded investigations are either Instrument based or are Interdisciplinary. Each of these are located at a Science Computing Facility (SCF). Using the EOS Investigators spreadsheet dated 20 April, 1993 (distributed by SPSO), the number of PIs and Co-PIs was identified. The number of support staff they would employ who would also make use of EOSDIS (e.g., research associates) was assumed to be, on average, three for each PI and Co-PI.

Table 2-1. Science Community Size

United States	
NASA EOS Funded Investigators	1900-3200
Instrument	(960-1600)
Interdisciplinary	(940-1600)
General Science	4200-8400
Academia	(3700-7300)
Federal Laboratories	(400-800)
Private Laboratories	(100-300)
U.S. Total	**6100-11600**
Other Countries	
EOS Investigators	280-470
Instrument	(40-60)
Interdisciplinary	(240-410)
Other Investigators	TBD*
Other Countries Total	**TBD**

* WAG of 4000-6000 International (General) Science Users was used as an estimate for SDR.

Methods for estimating the General Science community involved dividing the General Science category into three major sub-categories: federal laboratories and private laboratories conducting earth science research, and academia. Information for academic scientists in the United States was derived from *Peterson's Guide to Graduate Programs in the Physical Sciences and Mathematics, 1994*. The proportion of faculty to students was determined for a sample of University departments within the categories of Earth Science, Marine Science/Oceanography and Meteorology/Atmospheric Science. This proportion (21%) was then applied to today's number of departments within these categories to obtain an estimate of the potential user community (2500) for these disciplines. Since these categories in *Peterson's Guide* did not include all the potential users, the active membership (5.4%) or 1200 in three professional societies (IEEE Geoscience and Remote Sensing, Ecological Society of America, and the American Society of Agronomy) which may not be captured in the above estimates were added to the above numbers to arrive at the lower bound estimate. The number was doubled to obtain the upper limit.

In addition, an independent method of estimation was used to help verify the General Science estimates. A rapid literature survey was performed in which research articles from major earth science journals (*Journal of Geophysical Research - Atmospheres*, 1990; the *American Geophysical Union-Water Resources Research*, 1990; *Journal of Geophysical Research - Oceans*, 1990; *IEEE Geoscience & Remote Sensing*, 1990; and the *International Journal of Remote Sensing*, 1990) were categorized according to the geographical scale of the investigation and system interface type so as to map the results in to the science scenario matrix (described further in Technical Paper # 194-00313TPW, "User Characterization Methodology and Results"). The estimate (13,470) fell within the estimated range shown in Table 2-1.

The size of the Federal science community was obtained from questionnaires returned by Federal organizations. However, the number of General Science users does not include those communities who would be most interested in higher level investigation-derived products and other results which are yet to be defined. The estimated total size of the science user community is given in Table 2-1.

The size of EOS science community from other countries was derived in the same way as for the EOS Science category in the United States. The estimates for the General Science investigators from other countries are very preliminary and were based upon a survey conducted in Europe for European users of Earth observation data; estimates for other areas of the world are no more than subjective estimates by the contributors to this White Paper. If the size of this community is deemed to be important, more work is required to arrive at a defensible estimate.

2.1.2 Non-Science Users

Six major categories were used to sub-classify the Non-Science community: Federal government; state government; commercial end users, intermediaries, and education suppliers; education, primarily K-12 teachers and students; libraries; and policy makers. These non-science users are grouped together for much of the high-level analysis. Table 2-2 summarizes the projected size of the non-science community.

Table 2-2. Non-Science Community Size

(1999-2003 Time Frame)

United States	
Federal Government	1,500-2,200
States	1,500-3,000
Commercial-End Users	100-200
Commercial-Intermediaries	250-350
Commercial - Education Suppliers	80-140
Education (K-12) Teachers	2,000-7,000
Education (K-12) Students	58,000-174,000
Libraries	6,000-12,000
Policy Makers	TBD
Total	70,000-200,000

Questionnaires, interviews, and literature research were used to estimate the size of the non-science community and the scope of its potential use of EOSDIS. (See Technical Paper # 194-00313TPW, “User Characterization Methodology and Results” for more details on the methodology.) One of the more important assumptions is that the data products available through EOSDIS would be the standard data products designed for the needs of the science community. A judgment was made about the degree to which each of these products would be of direct interest to typical non-science users in each sub category in the 1999-2003 time frame. As the time horizon is extended beyond this period, it is expected that the size of the Federal, State and Commercial user communities will grow significantly as application techniques and models are developed to enable these communities to apply the EOS data products to their specific needs.

For high resolution sensors such as Landsat ETM and ASTER the demand is projected to be the highest for lower level (1B) data since previous experience indicates that non-science users, such as Federal and state agencies, and commercial organizations, would want to apply different algorithms than those of interest to scientists. For non-image data the demand would be greater for Level 2 and Level 3 products. Currently, the data product list (dated 14 February 1994) identifies very few Level 4 products, and does not address the data products that will be produced as part of the scientific investigations. Once these products are identified and the number of Level 4 products increases, it is anticipated that the estimated number of non-science users will increase significantly.

No estimates have been made as yet of the number of non-science users from other countries. It is expected that the use of EOSDIS by government and commercial organizations throughout the world will be much greater than the use by the science communities in other countries. The demand will expand as Internet and international commercial on-line services expand their global reach.

2.1.2.1 Federal Government

A questionnaire, based upon summary descriptions of each the standard data products (over 250 in number) that are to be available through EOSDIS, was developed and distributed within several major federal government organizations (DoI, DoE, NOAA, and USDA) and existing data centers (the DAACs and NOAA data centers). Twenty surveys were returned accounting for 70% of the projected users. Interviews (by telephone and in person) were conducted with selected government agencies and data centers to gain additional insight into their responses to the questionnaire.

Based upon data taken from questionnaires, responses to interview questions, and prior experience of the contributors to this White Paper, estimates are made for the size of the user community within other Federal organizations that are not associated with global change research. (Those associated with "global change" research are included under the "general science" category.) The estimates of the non-science users from Federal departments and agencies are as follows:

DoI	300-400
NOAA	600-700
DoE	300-400
USDA	100-200
EPA	20-100
DoD	150-350
Other	30- 50
Total	1500-2200 Users

2.1.2.2 States

Organizations, such as Departments of Natural Resources, Agriculture, and Environmental Quality, within four states (Texas, Alaska, Ohio, and Maryland) were selected for interviews to obtain first hand estimates of potential interest of state agencies for data products that will be available from EOSDIS. Two of these states, Ohio and Alaska, also completed the questionnaire, although it was designed for federal organizations. Based upon these results, estimates for the number of users in the applicable state were obtained. The remainder of the states were classified into High, Medium, or Low usage categories based on information from interviews and prior experience of contributors. The data obtained from questionnaires and interviews were then used to estimate the number of users in each category. The total size of the state user community follows.

High	50-100 Users(x17 states)
Medium	30- 60 Users(x15 states)
Low	10- 20 Users(x18 states)
Total	1500 - 3000 Users

2.1.2.3 Commercial Users

Commercial users include the "end user" companies with the capability to make direct use of the EOSDIS science data products to support business operations and various planning activities. Also included in this category are commercial intermediaries serving organizations that do not have in-house resources for data processing and analysis. Since EOSDIS data products are designed for use in scientific research, commercial users will rarely find direct and immediate application without additional processing and analysis of the data. Consequently the "end user" category includes only those companies that have the interest and resources to tailor products to their needs. This includes companies such as utilities, energy exploration and production companies, agribusiness, and major manufacturers and processors.

Based upon the experience of the contributors to this White Paper with previous remotely-sensed research products and upon interviews with NASA’s Centers for the Commercial Development of Space, we estimate that this community will be small, particularly in the early phases of EOSDIS, until the necessary R&D has been completed to develop techniques to apply the data to non-science applications. The estimates of 100-200 commercial end users and 250-350 commercial intermediaries for the 1998 time frame are based upon discussions with individuals at EROS Data Center (EDC), The Space Remote Sensing Center at the Stennis Space Center, and the Center for Mapping at The Ohio State University, all of whom have had direct experience in working with these communities.

The numbers of commercial education intermediaries were estimated with the assistance of data contained in the document: “Media Producers of CDROM/ Videodiscs” from the National Science Teachers Association of Science Education Suppliers, (1993). Their marketing information helped to determine whether the company was considered to be "large" or "small". The estimated number of users in each of the categories is as follows:

Large	3-5 (x15 companies)	45 - 75
Small	1-2 (x32 companies)	32 - 64
Total		80-140 users

2.1.2.4 Education

Currently the "educational community" only includes K-12 users; the undergraduate and graduate student population that is not part of the staff of a research scientist is not included. Additional research is needed to quantify the remaining college level user population.

The number of K-12 teachers in specialized science and social science fields such as Earth Science, Environmental Science, General Science, Physical Science, Physics, and Geography who would be interested in EOSDIS products is estimated to be 53,000 [1] in 1992. In 1998, the number is estimated to be 56,000 based on a 6% growth[2] in community size. Ninety percent (50,400)[3] of these teachers are expected to use microcomputers and networks. Based upon Texas Instrument studies and input from another vendor, we estimate only 5-15% of these teachers (2,520 - 7,560) will use these new "technologies" in the classroom for teaching purposes.

To estimate the size of the K-12 **student** user community, the figure of 23 students per teacher on average [4] was used resulting in 58,000-174,000 total potential EOSDIS student users.

To understand the magnitude of the student population the following are the counts for 1991 [5]:

Elementary students	26,425,503
Secondary students	20,262,769
Total K-12	46,688,272

The 58,000-174,000 projected number of K-12 student users is a very small percentage (0.3%-0.8%) of the total number of secondary students. If the current Administration (e.g., the Vice

[1]Market Data Retrieval, 1992.

[2]U.S. National Center for Education Statistics, Projections of Educational Statistics, biennial.

[3]U.S. Department of Commerce. Bureau of Census, Current Population Survey, Oct. 1984 and 1989. Unpublished data.

[4]National Science Teachers Association Reports, 1993.

[5]Market Data Retrieval, 1992. (Counts represent projected district-wide enrollments for Fall 1991, which were provided to MDR by each school district office. These counts do not represent the sum of school building enrollments.)

President and the NASA Administrator) follows-up its stated strong interest in expanding the use of on-line services and database availability to the education community with programs and funds, these projections will be low since the estimate is based on limited use of new technologies. Members of Congress and the Secretary of Education are proposing "---a federal policy that would ensure that schools are not bypassed as cable and telephone lines are installed for the electronic highway." [It is interesting to note that the use of Internet as of August 1993 by the K-12 community included a total of 111,000 addresses, as estimated by Tony Rutkowski when he was with Sprint International. However, how this community makes use of Internet, whether it is for use of e-mail or accessing data bases, is unknown.]

2.1.2.5 Libraries

The estimated number of libraries that will have the capability and interest in accessing EOSDIS is based upon statistics contained in the document: Library Mailing Lists 1992-1993, p.22 Market Data Retrieval 1992-93. There are 9,454 Main Public Libraries, and 6,648 Branch Public Libraries for a total of 16,202. However, it is assumed that only those with a book budget of over $20K, approximately 6,200 and 3,400 respectively, would better estimate those libraries that may provide access to EOSDIS for the communities they serve. When the estimated number of college and university libraries are included, the range of users associated with libraries who may make use of EOSDIS is between 6,000 and 12,000 users. Table 2-2 shows the number of library users within the context of the non-science ECS user community.

In a recent editorial in the *Washington Post*, August 1, 1994, Hardy R. Franlin, past president of the American Library Association, put forth the argument for connecting every public library to the "information superhighway" to give all citizens access to valuable databases. If this goal becomes a reality, our estimates could be low.

2.1.3 Service Providers

The estimates of the non-science community just discussed assume direct access to EOSDIS. However, another alternative for non-science usage of EOSDIS exists. As the demand from the non-science communities increases, other organizations may step forward to assist NASA in serving certain markets. For example, the Department of Interior's EDC, some of the NOAA data centers or commercial organizations may opt to assist in serving the needs of federal organizations, state agencies, commercial end-users, and intermediaries. NSF, NASA’s Office of Education or commercial enterprises may choose to establish a service tailored to the needs of the K-12 community. Similarly, commercial, county, or state organizations may decide to support the needs of librarians and the communities they serve.

Therefore there may be a select number of service/value-added data providers that interact with EOSDIS to provide a product more tailored to the needs of specific non-science communities. In the 1999-2003 time frame, the number of potential other service providers (organizations) are represented in Figure 2-1 by the numbers in parentheses. Generally the flow of data products will be from EOSDIS to these servers; however, Federal and state organizations may produce data products of value to NASA's science investigators. Consequently, there will probably also be a flow of data from Federal and state agencies into EOSDIS.

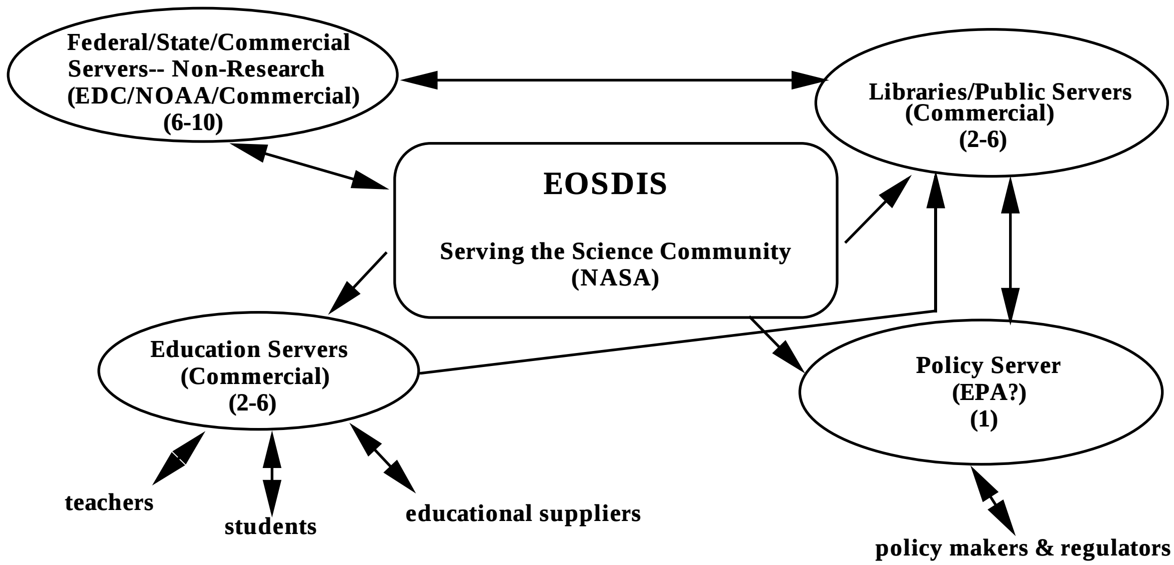


Figure 2-1 Service Providers

The nature of the demand on EOSDIS from the value-added providers will be strikingly different than if individuals from these communities entered EOSDIS directly. The value-added providers will tend to make greater use of machine-to-machine interfaces with EOSDIS and conduct business on a standing order basis, tailoring the EOSDIS products to the needs of various non-science communities, and making use of non-EOSDIS data as necessary to meet customer needs.

The number of routine EOSDIS users from these service providers can be reduced to 100-300 users rather than the expected 70,000-200,000 from the community at large. These numbers were derived by projecting the number of service providers in the 1999-2003 time frame and the number of user service personnel associated with each provider would be accessing EOSDIS to obtain products needed to serve their constituents. These assumptions are:

	Providers	User Service Personnel for each Provider
Federal/State/Commercial	6-10	10-15
Education	2-6	10-15
Library/Public	2-6	10-15

The number of users will depend on the nature of services provided by the other servers. For instance, if commercial service providers were to supply only an enhanced interface to EOSDIS and no valued-added products, the size of the community accessing EOSDIS directly would most probably be <u>greater than</u> 70,000-200,000. However, in this analysis, we will assume a lower number of users based on the assumption that these providers do not simply develop an interface to EOSDIS, but provide data products tailored to meet the needs of the individual non-science communities.

3. System Access

3.1 Frequency of Access

In the context of this paper the term “access” means a system level entry achieved by a user connecting his client software to the Data Server, or Advertising Service (for further information see SDS Section 4.5.2.3.1). This section describes frequencies and methods of system accesses by the user communities described in Section 2, and the services they would make use of in the course of accessing EOSDIS data products.

3.1.1 Science Community

To estimate the frequency of access for the science community, a classification was made based upon the number of accesses: yearly = 1-2, quarterly = 3-11, monthly = 12-24, weekly = 25-100, and daily = 100-250. Each science user scenario (see Technical Paper # 194-00313TPW, “User Characterization Methodology and Results”) was analyzed to determine how many accesses to the system would be made to complete the scenario. This was coupled with the demographic estimates in order to determine how often the entire science user community would access the system. For each scenario, the total accesses were mapped into the frequency classification. Table 3-1 summarizes these results. For the maximum user demographics there will be approximately an average of 47 accesses/user/year.

Table 3-1. Frequency of Access for Science Users

	Minimum	Maximum	Percentage
Yearly (1-2)	800	1500	13%
Quarterly (3-11)	1850	3500	30%
Monthly (12-24)	850	1600	14%
Weekly (25-100)	1800	3400	29%
Daily (100-250)	800	1600	14%
Note: Numbers in () indicate number of accesses/year			

3.1.2 Non-Science Communities

For each sector of the non-science community an estimate was made of the yearly demand for each of the data products (over 250 in number), including the demand for only browse products. This estimate was based on a questionnaire of Federal agencies and responses given in interviews of representatives from other non-science communities. Given these estimates and the total number of users in each sector (See Section 2.1.2), the frequency and number of accesses are calculated and presented in Table 3-2. This yields an upper limit of the total number of accesses of 640,000/year or an average of 3.2 accesses/user/year for the non-science community.

Table 3-2. Frequency of Access for Non-Science Users

	Minimum	Maximum	Percentage
Yearly (1-2)	65,200	186,200	93%
Quarterly (3-11)	3,700	10,600	5%
Monthly (12-24)	300	800	<1%
Weekly (25-100)	500	1,400	<1%
Daily (100-250)	300	1,000	<1%
Note: Numbers in () indicate number of accesses/year			

3.2 Access Methods

The methods that will be used for accessing ECS will not vary to any great extent with the user community. The trend is definitely towards the use of on-line or electronic access and, where routine access to large quantities of data is desired, direct machine-to-machine transfer will be employed. Also, since some components of the user community will be associated with other data and information systems, they would access ECS through these systems.

Access methods are important because they define what the various load accesses are and the loads on the system. They also help define what services are needed to support various modes of access, and provide insight into user environments. The fact that other data centers and individuals from other countries will be accessing EOSDIS indicates that services need to serve heterogeneous communities, and services that allow users to access EOSDIS through other systems need to be provided.

3.2.1 Access Means

The implicit analysis done when creating the scenario matrix was used to determine access means. Detailed science scenarios gave the breakdown found in Table 3-3. The demographics associated with a matrix cell for each category were added together to get the total number of users and percentage shown in Table 3-3.

Table 3-3. Access Means (U.S.)

Method	Percentage	Number(Users)
Telephone Interface Only	1.5%	90-170
Electronic	91.5%	5,600-10-600
Standing Orders	*Not*	*1,400-2,700*
Browsers	*Mutually*	*2,00-3,800*
Remote File Access (RFA)	*Exclusive*	*2,800-5,300*
Data Producers		*800-1,500*
Machine-to-Machine	7%	430-810
Total	**100%**	**6,100-11,600**

The means of access are divided into three major categories: telephone interface only, electronic, and machine-to-machine interface. The number and percentage of users that would use these different access methods (Table 3-3) is approximated using information collected in scenarios and questionnaires. The electronic category is further divided into four subcategories: 1.) standing orders, 2.) browsers , 3.) remote file access (RFA) users, and 4.) data producers. Standing orders are made by users who place orders for the future and who could be ordering for an extended period of time, for example, someone ordering a certain product, for a certain location, every ten days for the next year. Browsers are users that browse the data and do not do any other analysis on the system. Remote File Access users do some form of analysis on the EOSDIS system. They will also be using the computing power of the system. Data Producers are those users who are using data to produce different, and higher-level, data products that will be managed by ECS.

3.2.2 Entry Through Other Systems

Entry into EOSDIS will be in one of two ways: directly or through another data and information system. NOAA and other organizations in the U.S., Europe, and Japan will have somewhat similar environmental or Earth science data and information systems in the 1998-2003 time frame. The clientele that these systems are being designed to serve will have the option of entering EOSDIS via their own system, or entering EOSDIS directly. In general we expect that users would enter EOSDIS directly, unless their sponsoring organization, e.g., NOAA, DoI, Japan (NASDA/MITI), Europe (ESA, EC) paid (or made other arrangements for) whatever fees NASA would charge for EOSDIS data and services. In these cases they would most likely go through their sponsoring organization's data system. Estimates of the number of users entering EOSDIS directly and through other government data systems are given in Tables 3-4 (Science Users) and 3-5 (Non-Science Users). These estimates were also based upon discussions with individuals involved with the development and operations of data and information systems in DoI, NOAA, and DoE, as well as prior experience in working with people associated with European and Japanese data systems.

Table 3-4. Science Users Accessing Directly and Through Other Systems

United States	
	Numbers of Users
Direct Access	5,900-11,000
Access Through Other Data Systems:	
NOAA	100-300
Other	100-300
Total	**6,100-11,600**
From Other Countries	
Direct Access	2,900-4,500
Access Through Other Data Systems (e.g. Europe, Japan)	1,400-2,000
Total	**4,300-6,500**

Table 3-5. Non-Science Users Accessing Directly and Through Other Systems

United States Only (assumes NASA will serve these communities)	
Direct Access	69,000-198,000
Access Through Other Data Systems	1,000 - 2,000
(NOAA, DoI, Other)	
Total	**70,000-200,000**

3.3 Access Paths

Access paths and number of users provide sizing information for various system components. The relative use of the various services in accessing EOSDIS data is dependent upon the degree to which users in each sector are:

- familiar with EOSDIS data sets,
- familiar with the location of desired data sets,
- receiving their data through standing order,
- familiar with EOSDIS services,
- likely to search multiple datasets simultaneously crossing DAACs and/or SCFs, and
- exploring data sets and results outside their normal discipline.

The diagram below (Figure 3-1) illustrates the various pathways to the data server including the use of intermediary services (Advertising Service, Distributed Information Manager, Local Information Manager) to assist in locating data products. It is expected that as a user's familiarity with the system increases, his or her pathway through EOSDIS will change to direct access to data servers.

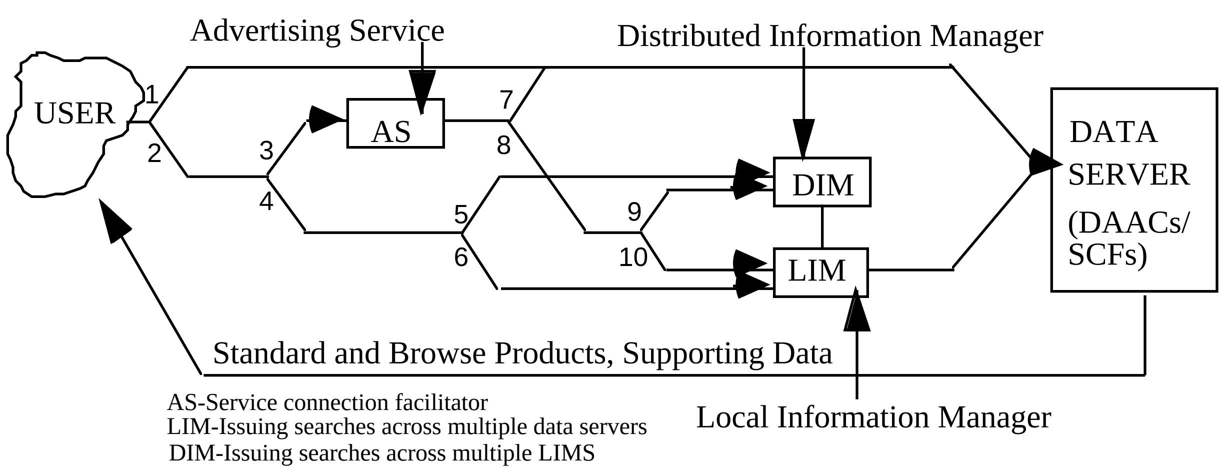


Figure 3-1. Data Server Access Routes

3.3.1 Science Community

To estimate the number of first accesses to different service components (Data Servers, Advertising Service, Distributed Information Manager, Local Information Manager) an analysis of the scenarios was conducted to determine how many accesses went to the DSs, AS, DIM and LIMs. Each scenario was analyzed to determine how many accesses to the system would be made to complete the scenario. Knowing the definition of the AS, DIM, LIM, and DS, an educated judgment was made as to which of these services was being utilized first in each access to the system. Once all the scenarios were completed, each of the service accesses was summed to get the percentages presented below for the data server access routes presented in Figure 3.1 and in Table 3-6. These first accesses to the system were summed to obtain the total number of accesses. The implications for the science community are that attention to services allowing direct use of data servers should be provided to support the science community.

The following numbers represent an estimate of the percentages for the accesses to the paths that correspond to Figure 3-1.

Path		
1	.867	(86.7% of users will access data servers directly)
2	.133	
3	.08	(8% of users will make use of the Advertising service)
4	.053	
5	.016	(1.8% of users, #5 + #9, will make use of the DIM)
9	.002	
6	.037	(4.5% of users, #6 + #10, will make use of the LIM)
10	.008	
7	.07	
8	.01	

3.3.2 Non-Science Community

For each sector of the non-science community an estimate was made, based upon the experience of the contributors to this White Paper in working with these communities, as to the percentage of users who would follow each of the ten paths identified in Figure 3-1. Attention should be paid to the efficiency of the DIM and LIM in order to support the non-science community numbers. When these Sector estimates were aggregated to the total non-science community, the following percentages resulted:

Path		
1	.04	(4% of users will access data servers directly)
2	.96	
3	.86	(86% of users will make use of the Advertising service)
4	.10	
5	.05	(9% of users, #5+#9, will make use of the DIM)
9	.04	
6	.05	(9% of users, #6+#10, will make use of the LIM)
10	.04	
7	.78	
8	.08	

Using the upper limit, the estimated total accesses per day (based upon a 250 working-day year) are given in Table 3-6.

Table 3-6. Access Paths to Data

Accesses/Day		
	Science	**Non-Science**
Direct to Data Servers	6,000	500
Use of Advertising Service	600	11,200
Through the LIM	200	1,200
Though the DIM	100	1,200

3.3.3 Implications

The majority of the science community will access data servers directly. The primary users of the Advertising Service, the DIM and the LIM will be the non-science communities and scientists who are seeking data in areas other than their normal discipline. Therefore, these services need to be designed to communicate with a large number of very diverse individuals .

Because of the potentially large number of non-science users, attention should be paid to the management of resources to allocate ECS services on a priority basis and to encouraging other value-added service providers to serve the non-science communities. This also implies that EOSDIS must be able to accommodate interaction with other service providers.

4. Data Extraction and Distribution

Volumes by DAAC and community, fraction distributed on physical media and by electronic transfer, and fraction of distribution via standing order and ad hoc request are just some of the information that will be discussed in the following chapter on data extraction and distribution. These are important in order to size components and to understand the lapse time between request and “pickup” required for sizing storage. Volume is important because it impacts storage loading, determines I/O and computing requirements, and communication bandwidth. The implications from the volume distributed versus the volume staged shows the amount of subsetting (~50%). Also, it implies different subsetting needs with a variability from 1/1 to 1000/1 when looking at the volumes staged to volume distributed.

4.1 Interest in the Standard Data Products

The important factors in the determination of the volume of data to be extracted and distributed are the relative interest the various user communities will have for the available data products, and the size of the associated communities. Different methodologies are used to estimate the size of the projected demand for the science and non-science communities.

4.1.1 Science Community’s Relative Discipline Focus

The relative interest in the standard data products from the science community was determined, in part, by the relative discipline focus of the EOS and General Science communities. The following two tables (Table 4-1 and 4-2) describe the relative product interest by discipline and by instrument and were obtained from an analysis of the Interdisciplinary Science investigator (IDS) lists from SPSO. The Table 4-3 is taken from an analysis of a literature survey of journals and the assignment of articles to the appropriate discipline. Details of this survey are described in the Technical Paper # 194-00313TPW, “User Characterization Methodology and Results”.

Table 4-1. Relative Product Interest, by Discipline, for EOS IDS Investigators

Atmosphere	44%
Land	32%
Ocean	14%
Cryosphere	6%
General	3%
Miscellaneous	1%

Table 4-2. Relative Product Interest, by Instrument, for EOS IDS Investigators

MODIS	55%
CERES	23%
ASTER	10%
MISR	7%
MOPITT	3%
LIS	2%

Table 4-3. Relative Sizes of the Science User Disciplines

Atmosphere	53.5%
Land	26.7%
Ocean	13.3%
Cryosphere	6.7%

The sources used in gathering this discipline information included the EOS Investigators list dated 20 April 1994, and the Peterson's Guide to Graduate Programs in the Physical Science and Mathematics, 1993.

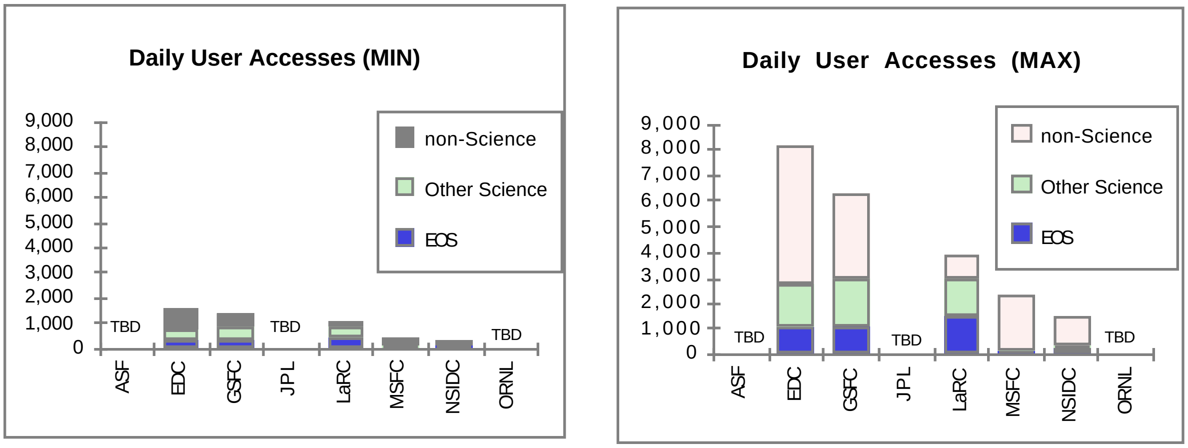
4.1.2 Non-Science Community

The relative interest in the standard data products by the non-science community was based upon responses to the questionnaires, interviews, and previous experience of the contributors to this White Paper in working with many elements of this community. Judgments were made as to the interest of each of the eight non-science sectors (listed in Table 2-2) for each of the 250+ standard data products, the geographic scale of interest for each product, and the number of times per year users would access each product and geographic scale. These estimates were then used to determine the number of user accesses for each layer of the data pyramid (See Figure 4-2 and 4-4) and the volume of data that would be accessed. The results are summarized in subsequent sections, 4.2 and 4.5.

4.2 User Accesses

The number of user accesses per day for each DAAC is given in Figure 4-1 and the relative interest by users in each of the data pyramid layers is given in Figure 4-2. The results are based upon analyses of the science scenarios, the returned questionnaires from the Federal agencies, discussions with state organizations, and interviews with educators, together with the derived results regarding the size of the potential user communities (Tables 2-1 and 2-2) and the relative interest in the standard data products as described in Section 4.1.

As expected, Figure 4-2 suggests that the science community will spend more time with the actual data than will the non-science community. The non-science users will make more use of the upper layers of the pyramid to obtain descriptive information about the data. The scientists, especially the EOS-funded scientists, will have more knowledge of the data in EOSDIS and what they require, and therefore will make somewhat fewer accesses to the upper pyramid layers. In both cases the results confirm the need for rapid access to the upper layers. Also since the access profiles across DAACs are rather heterogeneous, different design solutions are called for with different resource requirements.



Total = 4,478 user accesses/day

Total = 22,084 user accesses/day

Figure 4-1. User Accesses by DAAC

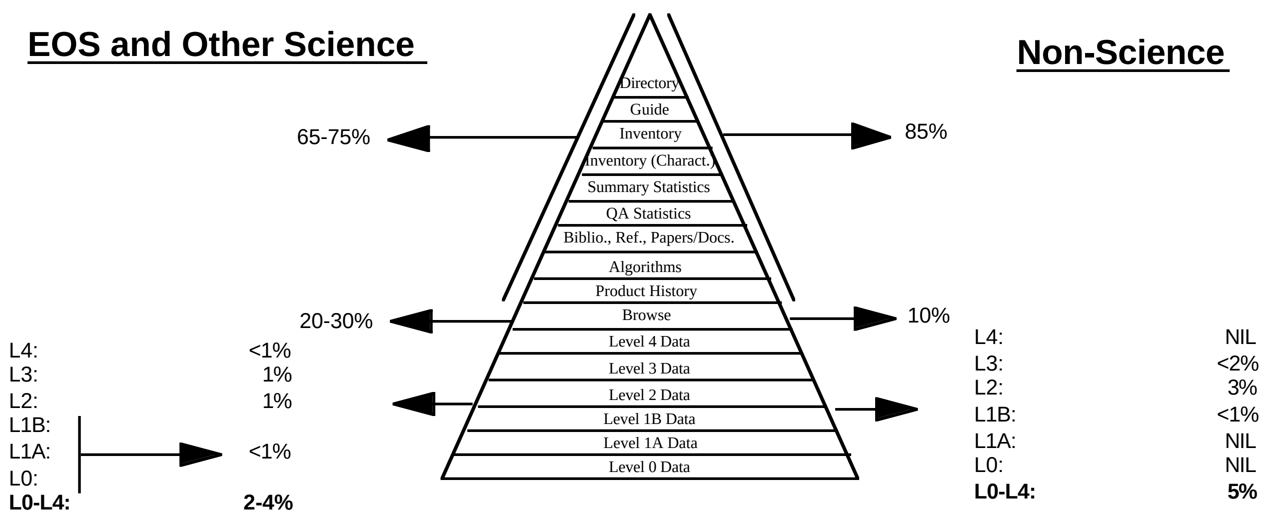


Figure 4-2. User Accesses by Pyramid Layer

The accesses by DAAC were derived from several analyses. First, the scenarios were analyzed to determine how many system accesses would occur in each scenario. The number of system accesses was then added up for all scenarios and then divided by the total users for the scenarios to get an average number of system accesses per user per year, which was 47 accesses for science users. This average was used with the Relative Product Access Frequency (RPAF) to get the distribution of accesses by DAACs.

The reason for the low projected demand for the L-4 products by the non-science community is due to the fact that very few such products are currently included in the Standard Data Product lists. As more descriptive information regarding L-4 products becomes available, it is expected that the projected demand will increase significantly. Also, while the percentages of accesses of browse products represents only 10% of the total accesses by this community, the number of browse products to be accessed per year could be in excess of 11,000,000 with the primary demand coming from the educational community. However, there is currently very little information available regarding the characterization of browse products to be produced, by whom, and when. Once this information is available, a more accurate assessment of the demand for browse products will be possible.

4.3 Types and Frequencies of Searches

A detailed understanding of how users will interact with the data and the system is difficult to obtain without an in-depth understanding of the process a scientist performs to complete her research. However, scenarios provided the details and understanding necessary to examine the search service. Because the search service was determined to be one of the more important design drivers, this section singles out search accesses from all other accesses. The characterization of user interaction with other services will be developed in support of PDR.

The scenarios were analyzed to find the percentage of users and the frequency with which they would be conducting simple searches, content searches, or coincident searches. Each scenario collected user services, and search was one of these services. The searches for all the scenarios were reanalyzed and categorized into one of these three types of searches. The later of these two types of searches have subcategories. Content searches are divided into subset and data content searches. Coincident searches are divided into user refined, match-up, and complex coincident searches. The estimated percentages are found in Table 4-4.

Table 4-4. Percentage of Different Search Types

	Definitions	**% of Total Searches**
Simple Searches	Include all searches not defined below	60-75%
Content Searches		10-15%
Subset	Searching for spatial/temporal scales, or parameters	
Data Content	Searching on certain data characteristics	
Coincident Searches		15-25%
User Refined Coincident Searches	User makes two separate queries	
Match-up Coincident Searches	User uses one table to get a match from a second table	
Complex Coincident Searches	i.e. LIS, over Texas, in June 1997, where NLDN Lighting Data is available	
*Note: Not all accesses are Searches.		

The dynamic nature of this capability, and the leading-edge technology requirements for these kinds of complex search modes, implies that more research needs to be conducted in this area to obtain a more complete understanding of users' search patterns and complexities

4.4 Geographic Scale of Interest

For a given data product, the volume of data accessed by users is directly related to the geographic scale of their interest. Geographic scales of interest were estimated in different ways for the science and non-science users. The abscissa of the science scenario matrix represents the geographic scales of interest of the science community. [Users were divided into four categories which correspond to geographic scale of interest with the following titles and ranges: Review (sample data), Local/Field/Case Studies (<1000 km^2 or 400 mi^2), Regional Studies(>1000 km^2 but < 10^7 km^2 or 4 x 10^6 mi^2), Global Studies (>10^7 km^2 or 4 x 10^6 mi^2)]. Table 4-5 shows the number that was used to simplify the range for the calculations corresponding to the volumes staged and distributed. (See “User Characterization Methodology and Results” White Paper #194-00313TPW.) The non-science geographic scales were estimated using information obtained through questionnaires and through interviews with various communities. Questionnaires from the government agencies were analyzed to understand geographic scale and to understand size of data they were projecting to use from EOSDIS.

More detailed information should be collected in this area to obtain a more complete understanding of not only the users that will use different scales, but also the frequency of coverages that they would require.

Table 4-5. Geographic Scale of Interest

Science Users	Percentage	Number of Users per Year
Browse Products **Only**	NIL	NIL
$1x10^2$ KM^2	8%	500-900
$1x10^3$ KM^2	39%	2,400-4,500
$5x10^5$ KM^2	19%	1,200-2,200
$1x10^8$ KM^2	34%	2,000-4,000
Total	**100%**	**6,100-11,600**
Non-Science Users		
Browse Products **Only**	93%	14,500-186,800
$1x10^2$ KM^2	3%	2,600-7,200
$1x10^3$ KM^2	2%	1,800-4,300
$5x10^5$ KM^2	<2%	800-1,300
$1x10^8$ KM^2	<1%	300-400
Total	**100%**	**70,000-200,000**

4.5 Volume Staged and Distributed

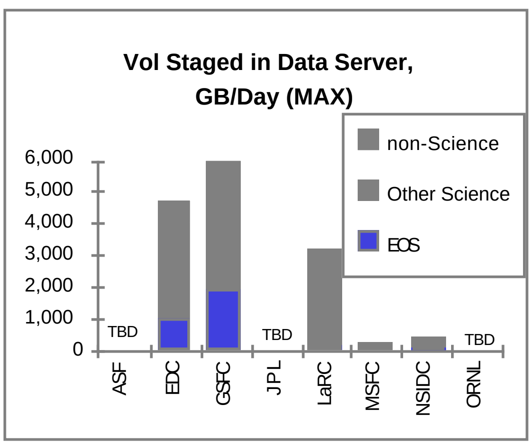
Volume is important because it impacts storage loading, determines I/O and computing requirements, and communication bandwidth. This can be important in order to size components and to understand the lapse time between request and “pickup” required for sizing storage. The implications from the volume distributed versus the volume staged shows the amount of subsetting (~50%). This also, implies that each community has different subsetting needs with a variability from 1/1 to 1000/1 when looking at the volume staged to volume distributed.

To obtain an estimate of the volume of data to be pulled by the user communities, summaries were prepared of the total demand in a year for each standard data product, for each geographical scale (i.e., 1x 10^2 km^2, $1x10^3$ km^2, $5x10^5$ km^2 and $1x10^8$ km^2) and for each sector of the user community (for more details see Technical Paper # 194-00313TPW, “User Characterization Methodology and Results”). With the size of each data granule and the required subsetting known, the volume, both staged and distributed to users, were calculated.

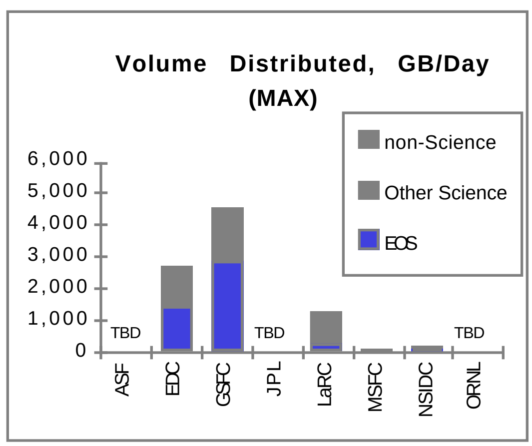
The volumes estimated are for the 1999-2003 time frame. It is assumed that by this time, direct parameter-level access is possible, both out of the archive and from the processing stream. Products, product sizes, and granule definitions are consistent with the SDR product baseline as of May 10, 1994. While estimates were made for the projected demand for Landsat and SAR data products, time did not permit the inclusion of this data at SDR. Since there is currently no adequate definition of browse products that will be available, the browse volumes are not included. However, there appears to be a strong potential demand for browse products, especially for the “other-science” and non-science communities.

The results are also based upon a 365 day per year system operation with a 250 day per year user work year. There is no latency assumption in “user retrieval” estimates (other than 365 vs. 250). Data volumes due to standing orders and ad-hoc requests were combined, as were volumes from electronic and media transfer. The availability of a “smart” subscription service was assumed where initial subscription enrollment includes user specification of desired geographic regions and parameters in which only the data whose granule boundaries satisfy those specifications are staged/distributed. EOS Users’ by-DAAC volume proportions were applied to “other science” users’ (same as General Science Users) total volumes (no RPAFs used).

The numbers for the IDS investigators (Table 4-6b) are worst case estimates developed by NASA and are expected to be significantly reduced in the future. Therefore, Figure 4-3, 4-5 and Table 4-6a are believed to overstate the projected volume distributed. Figure 4-3 gives the resulting bounding estimates for the total user pull requirements for data staged in the data servers, subsetting amount (implicit), and distribution. The estimates for each of the three components of the user community contributing to the total are also given (Table 4-6) for illustrative purposes.



Total = 14,393 GB/day
(5,939 GB/day - MIN)



Total = 8,550 GB/day
(3,991 GB/day - MIN)

Figure 4-3. Projected Data Volumes

Table 4-6a. Total Data Volumes Staged and Distributed

TOTALS	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	1,906,954	4,662.746	1,196,523	2,626,236
GSFC	2,626,803	5,964,737	2,259,538	4,507,033
JPL	TBD	TBD	TBD	TBD
LaRC	1,176,414	3,152,429	460,055	1,206,762
MSFC	81,907	196,092	15,970	48,210
NSIDC	146,932	417,348	58,952	162,134
ORNL	TBD	TBD	TBD	TBD
TOTAL	**5,939,010**	**14,393,351**	**3,991,038**	**8,550,374**

Table 4-6b. EOS IDS and Instrument Investigators' Data Volumes Staged and Distributed

EOS IDS +Instrument	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	522,826	980,713	763,849	1,432,821
GSFC	1,042,165	1,911,298	1,522,603	2,792,407
JPL	TBD	TBD	TBD	TBD
LaRC	91,523	151,432	133,715	221,242
MSFC	170	338	248	494
NSIDC	6,723	13,318	9,822	19,458
ORNL	TBD	TBD	TBD	TBD
TOTAL	**1,663,406**	**3,057,099**	**2,430,236**	**4,466,422**

Table 4-6c. Other (General) Science Data Volumes Staged and Distributed

Other (General) Science	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	984,378	2,989,236	355,482	1,079,483
GSFC	1,076,938	3,270,310	388,908	1,180,986
JPL	TBD	TBD	TBD	TBD
LaRC	894,986	2,717,781	323,200	981,454
MSFC	43,484	132,046	15,703	47,685
NSIDC	124,101	376,854	44,816	136,091
ORNL	TBD	TBD	TBD	TBD
TOTAL	**3,123,888**	**9,486,228**	**1,128,109**	**3,425,699**

Table 4-6d. Non-Science Data Volumes Staged and Distributed

Non-Science	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	399,750	692,797	77,193	113,932
GSFC	507,700	783,128	348,027	533,640
JPL	TBD	TBD	TBD	TBD
LaRC	189,905	283,216	3,139	4,066
MSFC	38,253	63,708	19	31
NSIDC	16,108	27,176	4,315	6,585
ORNL	TBD	TBD	TBD	TBD
TOTAL	**1,151,717**	**1,850,024**	**432,994**	**658,254**

Inspection of these tables shows that the MIN/MAX estimates span a factor of 2 to 3 (not including Landsat data). Relative to the production volume (2.1 TB/day), the volume staged in the data server due to user pull is greater by a factor of 2.8 (min) and 6.7 (max). The volume distributed to all users is greater than the production volume by a factor of 1.9 (min) and 4.0 (max). The subsetting ratio is 2-3:1 (on average, after correcting for 250 vs. 365 day years). This ratio is substantially higher at some DAACs and for some user groups: 10-1000:1 (See Figure 4-3). Smaller granules will reduce the amount of subsetting required and the volume staged. It is also important to note that the volume/request (~2 GB/request, not shown in table) is consistent with current Goddard DAAC experience with "EOS-like" data. Calculations were based upon the granule sizes contained in the SPSO data base, which has not been verified by the instrument teams.

Projected volume distributed by pyramid layer is shown in Figure 4-4 below. It is apparent that most of the volume of data distributed is from the lower levels of the data pyramid.

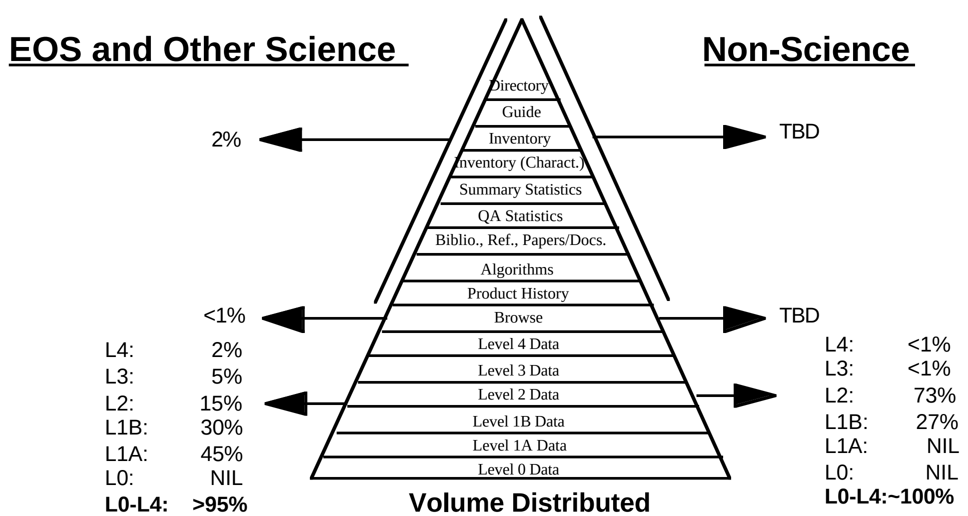


Figure 4-4. Projected Volume Distributed by Pyramid Layer

4.6 Data Distribution Media

The analysis of scenarios and questionnaires indicates that 15% of users will use some sort of physical media to receive their data, while 85% will require electronic transfer for the distribution of their data. However this does not take into account the cost of electronic data transfer at the price elasticity of demand, which are to be collected to support PDR. These same sources indicate that the access mode will be approximately 30% via standing orders and 70% via on-demand/ad hoc requests.

These results regarding distribution are just the beginning of the analysis that needs to be conducted in this area. For instance, the volume of data to be distributed on physical media and via electronic means needs to be assessed. Design must be able to handle a high number of electronic distributions, and potentially high volumes. The design must look at the question of automatic transfer vs. user initiated transfer. Since there is a potential for significant physical media demand (~5%) on standing orders, design must consider impacts on physical distribution services.

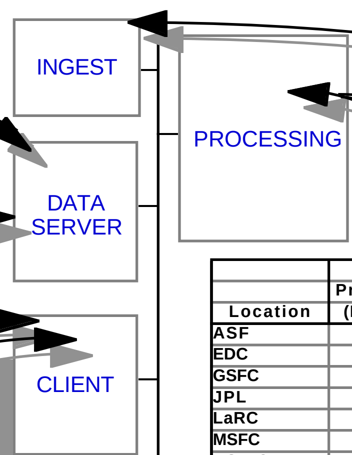
4.7 Results

The volume before subsetting, the number of daily user accesses and the total volume distributed to the users per day represent the maximum case. The derivations of these estimates are explained in previous sections. The peak processing, peak I/O and storage requirements are results from the Strings Model.

A summary of the estimated input, processing and output data volumes for each of the DAACs is given in Figure 4-5 (note: the volumes for EDC do not include Landsat data).

Location	L0 Volume (GB/Day)
ASF	
EDC	90
GSFC	67
JPL	
LaRC	41
MSFC	<1
NSIDC	
ORNL	
TOTAL	198

Location	Processing Output (GB/Day)
ASF	
EDC	1124
GSFC	974
JPL	
LaRC	370
MSFC	1
NSIDC	3
ORNL	
TOTAL	2473



Location	Peak Processing (MFLOPS)	Peak I/O (MB/sec)	Peak Storage (GB)
ASF			
EDC	4081	385	511
GSFC	20212	423	3687
JPL			
LaRC	14973	265	1043
MSFC	127	7	1
NSIDC	81	2	2
ORNL			
TOTAL	39474	1082	5244

Location	Volume Before Subsetting (GB)	Daily User Accesses	Volume Distributed (GB)
ASF	TBD	TBD	TBD
EDC	4663	8130	2626
GSFC	5965	6302	4507
JPL	TBD	TBD	TBD
LaRC	3152	3825	1207
MSFC	196	2286	48
NSIDC	417	1542	162
ORNL	TBD	TBD	TBD
TOTAL	14393	22085	8550

Figure 4-5. System Volume Summary

These results indicate that the estimated maximum volume to be distributed to the total user community will be approximately 3.5 times the daily production or processing output. The results also indicate that a volume of 14.3 TB/day will require subsetting to generate the distributed volume of 8.55 TB/day. To reduce the amount of processing required, methods should be explored that would permit reducing this amount of subsetting.

5. Observations and Implications

While the primary focus of this paper has been on the quantitative results, the following is a summary of the more important qualitative observations from this analysis and their general implications.

Observation The findings to-date indicate that approximately 85% of the users will want distribution of data by electronic means. This of course will be tempered by the cost of electronic data transfer.

Implication As network capacity grows and cost are reduced, ECS must accommodate a large number of users employing electronic transfers. To refine the estimated use of electronic data transfer, projections of the cost of such transfers in the 1998-2003 timeframe should be developed, and estimates should be made of the price elasticity of demand for the various user communities.

Observation Almost 30% of the science users will request distribution through standing orders.

Implication EOSDIS should provide efficient mechanisms for routine distribution.

Observation Although the EOSDIS data products are being designed to meet the needs of scientists, there are potentially a large number (70,000-200,000) of U.S. non-science users who will be interested in these products.

Implications ECS should plan to accommodate interaction with other service-providers that may choose to serve these markets with products and services based inpart on EOSDIS data products . In addition, the allocation of ECS services and resources needs to be managed to ensure that priority is given to the needs of the Earth science community.

Observation The same data is used by a diverse user community. Diversity of interest extends to scale and nature of queries.

Implications The EOSDIS system must provide views of data customized for different disciplines and communities. Access to data should also be provided at the parameter level.

Observation Many users will have difficulty dealing with information that is very product, discipline, and instrument specific.

Implication The design must bridge differences and support system-wide searches.

Observation   Users cannot be expected to know all aspects of metadata, including method of derivation and size.

Implications   Interface must provide visibility to some users of the inputs and processing used to produce the data products and transparency to others. ECS must provide mechanisms for informing users of the resources required to meet their needs and for controlling the use of the system's resources.

Observation   The complexity of searches varies from simple to complex, with a potentially large number of complex searches.

Implications   ECS must provide support for efficient searches of all types.

Observations   The total volume of data requested from each DAAC varies tremendously.

Implications   ECS Architecture must be highly scalable among DAAC storage and throughput, and local network configurations.

6. Current Status of Analysis

6.1 EOSDIS Data Inputs

The data inputs that were taken into account in producing the results contained in this report are given in Table 6-1. The user community demand and data volume for the Level 4 products and other investigator results that are to become available through EOSDIS need to be estimated and factored into the data ingest and processing requirements for the individual DAACs. In addition, the volume estimates for Landsat, SAR, and demand for V0 data products should also be estimated. Once the ancillary and correlative data requirements are better understood, the plan is to factor these requirements into the analysis.

Table 6-1. EOSDIS Data Inputs

Current Status of Analysis	
EOS Platforms	√
IDS Level 4 Products and Investigation Results	TBD
TRMM	Partial
Landsat 7	Partial
SAR Missions (E-ERS-2/J-ERS-1/Radarsat	Partial
Ancillary Data	TBD
Correlative Data	TBD
V0 migration/Pathfinders	TBD

6.2 Follow-On Analysis Required

Since it is impossible to predict the future, user characterization and enhancing our understanding of their ever-evolving performance requirements is a continuous process, involving targeted surveys, questionnaires, interviews, analyses, modeling and testing (with Versions 0, 1 etc.) to detect and discern the more detailed and evolving requirements. Other uncertainties in the assumptions, inputs, or limitations of this study that need to be addressed and factored into our understanding of the interaction of the user communities with EOSDIS include:

- the time-dependent aspects of user access and system utilization, and how this will evolve from Release A through Release D, including understanding search patterns that users will employ and their complexity;
- the effect of projected total usage cost (NASA fees, on-line service charges, the use of high capacity commercial cable systems, etc.) to users on projected demand from each user community sector;

- the insights that will be gained from Version 0 usage statistics and historical trends of other government environmental data centers;
- the study of alternative service scenarios for value-added resuppliers meeting at least a significant portion of the projected demand from the non-science community;
- an improved understanding of the General Earth Science community and how they will make use of EOSDIS;
- assessing whether the demand from the domestic under-graduate and foreign general science and non-science communities could be significant enough to warrant more-detailed investigation;
- definition of browse products that should be available to the various user communities and a reassessment of the user demand;
- an assessment of the volume of data to be distributed by physical media and electronically, taking into account cost and price elasticity of demand; and,
- an assessment of user interaction with all planned services.

A plan for conducting these and other tasks necessary to support PDR is being developed and will be issued as a separate document. The plan will be driven by the needs of SDPS and CSMS to support PDR.

Abbreviations and Acronyms

AS	Advertising Service
CIESIN	Consortium for International Earth Science Information
CSMS	Communications and System Management Segments
DAAC	Distributed Active Archive Center
DIM	Distributed Information Manager
DoE	Department of Energy
DoI	Department of the Interior
DS	Data Server
EC	European Community
ECS	EOSDIS Core System
EDC	EROS Data Center
EOSDIS	EOS Data and Information System
EPA	Environmental Protection Agency
EROS	Earth Resources Observation System
ESA	European Space Agency
IDS	Interdisciplinary Science
K-12	Kindergarten through 12th Grade
LIM	Local Information Manager
MITI	Ministry of International Trade and Industry
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency
NLDN	National Lightning Detection Network
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
PDR	Preliminary Design Review
PI	Principal Investigator
R &D	Research and Development
RFA	Remote File Access

RPAF	Relative Product Access Frequency
SCF	Science Computing Facility
SDPS	Science Data Processing Segment
SDR	System Design Review
SDS	System Design Specification
SPSO	Science Processing Support Office
USDA	U.S. Department of Agriculture